

Einführung in die digitale Langzeitarchivierung

Schumann, Natascha

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Schumann, N. (2012). Einführung in die digitale Langzeitarchivierung. In R. Altenhöner, & C. Oellers (Hrsg.), *Langzeitarchivierung von Forschungsdaten: Standards und disziplinspezifische Lösungen* (S. 39-50). Berlin: Scivero Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-457405>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-ND Lizenz (Namensnennung-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-ND Licence (Attribution-NoDerivatives). For more Information see:
<https://creativecommons.org/licenses/by-nd/4.0>

Einführung in die digitale Langzeitarchivierung

Natascha Schumann

In fast allen Bereichen des öffentlichen, kulturellen und wissenschaftlichen Lebens, einschließlich des privaten Bereichs ist eine Zunahme an der Erstellung und des Gebrauchs digitaler Objekte zu beobachten. Die Verbreitung und Vorhaltung der Daten in digitaler Form steigt nicht nur quantitativ an, sondern ersetzt in vielen Bereichen bisher genutzte analoge Formate.

Im Bereich der wissenschaftlichen Publikationen kann man diesen Wandel gut nachvollziehen. So besteht seit ungefähr 15 Jahren die Möglichkeit, Dissertationen und Habilitationen in elektronischer Form zu publizieren. Abgesehen von den finanziellen Vorteilen für die Autorinnen und Autoren, bietet diese Veröffentlichungsvariante zudem einen schnellen und weltweiten Zugriff auf aktuelle wissenschaftliche Ergebnisse. Mehr als 100.000 Online-Dissertationen aus dem deutschsprachigen Raum sind im Katalog der Deutschen Nationalbibliothek nachgewiesen. Darüber hinaus wächst die Anzahl der Netzpublikationen generell und der E-Books ebenfalls stetig an, so sind an der Deutschen Nationalbibliothek mittlerweile mehr als eine halbe Million elektronischer Publikationen (Online-Dissertation, Online-Artikel und E-Books) verzeichnet. Ebenso werden in zahlreichen Archiven Vorgänge zunehmend in elektronischen Akten verwaltet.

Im wissenschaftlichen Bereich sind elektronische Publikationen und Daten gar nicht mehr wegzudenken. Neben den genannten Publikationen entstehen enorme Datenmengen allein aus den unterschiedlichen Forschungsgebieten. Nicht nur in der Klimaforschung fallen enorme Klimadaten an, auch in weiteren naturwissenschaftlichen Fächern werden bei Experimenten große Datenmengen generiert. In den Sozial- und Wirtschaftswissenschaften werden große Mengen unterschiedlicher Forschungsdaten bei quantitativen und qualitativen Studien erzeugt.

Es entstehen also in allen Bereichen digitale Daten, die für unterschiedlich lange Zeiträume bewahrt werden müssen. Das stellt uns vor neue Herausforderungen, denn die bewährten Methoden der Bestandserhaltung für analoge Objekte können nicht einfach auf digitale Formate übertragen werden. Abgesehen von der grundsätzlich anderen Struktur digitaler Anwendungen, zum Beispiel in Bezug auf Verknüpfungsmöglichkeiten zu anderen Objekten, gibt es bei der Erhaltung digitaler Daten mehrere Dinge zu beachten: Digitale Daten sind an einen Datenträger gebunden. Diese unterliegen einem schnellen technologischen Wandel, man denke zum Beispiel an die verschiedenen Formate der letzten zwei

Jahrzehnte, von der Floppy-Disk zum USB-Stick. Werden Datenträgerformate obsolet, ändert sich entsprechend auch die Hardware, das bedeutet zum Beispiel, eine Diskette kann rein physisch nicht mehr gelesen werden, wenn der Rechner kein Diskettenlaufwerk mehr besitzt. Darüber hinaus ist es auch nötig, dass das Betriebssystem mit dem Datenträger umzugehen weiß und die entsprechende Software noch verfügbar ist. Ein weiteres Problem liegt darin, dass Datenträger über die Zeit beschädigt werden können und dann ebenfalls nicht mehr korrekt angezeigt werden können.

Bei der Erhaltung digitaler Daten spielen also verschiedene Faktoren eine Rolle, die jeweils miteinander verknüpft sind. Das bedeutet, verschiedene Komponenten entscheiden darüber, ob ein digitales Objekt in Zukunft nicht nur lesbar, sondern auch korrekt interpretierbar bleibt.

Die herkömmlichen Strategien zur Erhaltung schriftlichen Kulturguts sind über einen sehr langen Zeitraum entstanden und erprobt worden. So war in früheren Zeiten der Aufwand für die Erstellung eines gedruckten Buches größer als heute die Erstellung eines Textes in digitaler Form. Dafür war und ist die Nutzbarkeit doch relativ einfach zu gewährleisten. Werden Faktoren wie beispielsweise säurefreies Papier, die richtige Luftfeuchtigkeit und Raumtemperatur berücksichtigt, kann ein Buch über mehrere Jahrhunderte erhalten und interpretierbar bleiben, ohne dass das Objekt selbst verändert werden müsste. Bei digitalen Objekten hingegen sieht das etwas anders aus. Der Aufwand zur Erstellung eines digitalen Textes ist heutzutage relativ gering, die Sicherstellung der Verfügbarkeit und der richtigen Interpretierbarkeit ist jedoch sehr viel komplexer als bei gedruckten Medien. Es muss immer wieder überprüft werden, ob die digitalen Objekte mit gängigen technologischen Mitteln noch funktionstüchtig sind. Das setzt voraus, dass es Mechanismen geben muss, mit denen das (sich ständig weiterentwickelnde) Marktgeschehen beobachtet werden kann. Nur so ist es möglich zu entscheiden, welche Formate mittelfristig nicht weiter unterstützt werden, um dann entsprechende Maßnahmen zur Sicherung der Daten einleiten zu können.

Im Bereich der digitalen Langzeitarchivierung wird in der Regel auf das OAIS-Modell Bezug genommen. Das Open Archival Information System¹ gilt als Referenzmodell und wurde Ende der 1990er Jahre von der NASA in Zusammenarbeit mit verschiedenen Raumfahrtorganisationen entwickelt. Es handelt sich um ein sehr generisches Modell, welches verschiedene Komponenten der Langzeitarchivierung berücksichtigt. Dabei werden nicht nur technische Aspekte behandelt, sondern auch die organisatorischen Bedingungen einbezogen. Es wird nicht vorgegeben, wie die einzelnen Schritte in konkreten Anwendungen umgesetzt werden.

¹ <http://public.ccsds.org/publications/archive/650x0b1.pdf>

Deutsche Übersetzung: <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2012051101>

Die wesentlichen Aspekte des Modells bestehen darin, dass ein Produzent ein Objekt erstellt und es als SIP (=Submission Information Package) in das digitale Archiv abliefern. Der Vorgang der Übernahme in das Archiv wird Ingest genannt. Dieses Paket wird im Archiv mit weiteren Informationen angereichert, die im Hinblick auf die langfristige Verfügbarkeit des Objektes von Bedeutung sind. Dabei handelt es sich um verschiedene Metadaten, zum Beispiel solche, die Angaben zum Format samt Version oder zu technischen Aspekten enthalten. Das nun erweiterte Objekt wird AIP (Archival Information Package) genannt. Innerhalb des Archivs sind weitere Prozesse definiert, darunter die Verwaltung des Objektes, das Datenmanagement und die Archivierungs-/Speicherfunktionen. Als zusätzliche Aufgabe ist das Preservation Planning genannt. Damit ist ein Bündel von Maßnahmen gemeint, die es erlauben, die richtigen Entscheidungen im Hinblick auf die Erhaltung der digitalen Bestände zu treffen. Dazu zählen unter anderem die Beobachtung der technologischen Entwicklungen, zum Beispiel in Form von Technology Watch, das Sammeln und Pflegen von Informationen zu Formaten sowie zu Soft- und Hardware. Ebenfalls zu diesem Bereich gehört die Definition dessen, was für verschiedene Gruppen von Objekten die wichtigsten Eigenschaften (auch „Signifikante Eigenschaften“ genannt) sind, die es auf jeden Fall zu erhalten gilt. Für den Zugriff der Objekte durch Nutzer wird auf Grundlage des SIP ein gesondertes Paket generiert, das DIP (=Dissemination Information Package). Das ist im Regelfall die jeweils aktuellste Version eines Objektes, welches mit gängigen Systemen genutzt werden kann.

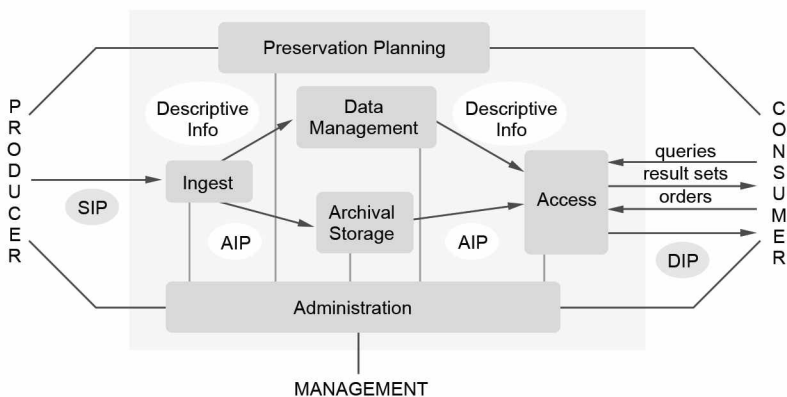


Abbildung 1: OAIS-Referenzmodell

Für ein digitales Archiv spielen die Frage der Authentizität und der Integrität von Objekten eine zentrale Rolle. Durch die Tatsache, dass zunehmend mehr Dokumente digital vorliegen, stellt sich die Frage, wie gewährleistet werden kann, dass diese Dokumente auch über einen längeren Zeitraum unverändert bleiben.

Das bedeutet, für digitale Dokumente müssen Kriterien eingehalten werden, die gewährleisten, dass Informationen und Dokumente echt sind, im Sinne, dass sie weder beschädigt noch manipuliert worden sind. Sowohl im Hinblick auf die Berechnung der Rente als auch beispielsweise in Bezug auf Grundstücksgrenzen bedarf es verlässlicher Daten, die auch zukünftig die Informationen wiedergeben, die sie bei der Erstellung enthalten haben. Darüber hinaus müssen diese Daten auch korrekt interpretierbar sein. Ein weiterer Punkt ist die Zugänglichkeit zu den Objekten, die durch geeignete Maßnahmen wie beispielsweise entsprechende Bereitstellungssysteme gewährleistet sein muss.

Um all diese Anforderungen erfüllen zu können, gibt es verschiedene Strategien und Methoden. Als Grundlage für alle weiteren Maßnahmen muss zunächst die physische Datensicherung erfolgen. Die Einsen und Nullen müssen erhalten bleiben. Diese Sicherung nennt sich Bitstream Preservation.

Digitalisierung wird teilweise als Erhaltungsstrategie bezeichnet. Dabei muss jedoch berücksichtigt werden, dass dies in der Regel im Zusammenhang mit analogen Objekten geschieht, die durch Digitalisierung erhalten werden sollen. Zum einen bietet dies den Vorteil, dass beispielsweise sehr wertvolle und von Zerfall bedrohte Werke so weiterhin Nutzern zugänglich gemacht werden können. Zum anderen können dadurch sehr viel mehr Nutzer erreicht werden. In Bezug auf digitale Langzeitarchivierung muss deutlich gemacht werden, dass Digitalisierung keine Erhaltungsstrategie darstellt. Denn Digitalisate bedürfen der gleichen Maßnahmen zur Erhaltung wie digital entstandene („digital born“) Objekte. Die derzeit am meisten genutzten Erhaltungsstrategien sind Migration und Emulation, die alternativ oder je nach digitaler Sammlung auch ergänzend eingesetzt werden.

Die vor allem bei Textdokumenten gängigste Methode, um diese über einen möglichst langen Zeitraum verfügbar zu halten, ist die Migration. Darunter versteht man die regelmäßige Konvertierung von Objekten in jeweils aktuelle Datenformate. Das kann sowohl bedeuten, dass in ein komplett anderes Format konvertiert wird, zum Beispiel von einem Worddokument in ein PDF. Aber eine Migration kann auch bedeuten, dass innerhalb eines Formats in eine neuere Version kopiert wird. Der Vorteil der Migration liegt darin, dass die digitalen Objekte immer in einem aktuell zugänglichen Format vorliegen, das mit gängiger Software lesbar ist. Der Nachteil liegt darin, dass eine Migration durchaus die Gefahr von Datenverlusten beinhalten kann. Darüber hinaus unterliegen auch die „neuen“ Formate, in die migriert wurde, einem Alterungsprozess. Das bedeutet, dass in regelmäßigen Abständen überprüft werden muss, ob das Format noch unterstützt wird und dann gegebenenfalls eine neue Migration in die Wege geleitet werden muss. Insgesamt wird es im Lebenszyklus eines digitalen Objektes mehrere Migrationen geben. Um eine Migration durchführen zu können, ist es notwendig, entsprechende Informationen über das vorliegende For-

mat zu besitzen. Daher werden bereits bei der Ablieferung in das digitale Archiv (Ingest-Prozess) Metadaten extrahiert und erfasst. Dafür stehen verschiedene Tools zur Verfügung, die diesen Prozess automatisch durchführen. Dadurch wird ermöglicht, dass bei einer anstehenden Migration alle Objekte eines bestimmten Formates mit der Version xy identifiziert und dann in das neue Format migriert werden können. Je komplexer ein Dateiformat ist, desto aufwändiger stellt sich die Migration dar. In einigen Fällen ist sie daher nicht das geeignete Mittel zur Erhaltung der Daten.

Eine weitere Strategie zur Langzeitarchivierung ist die Emulation. Hierbei wird nicht das digitale Objekt selbst verändert, wie bei der Migration, sondern es wird eine Umgebung geschaffen, die das Objekt in adäquater Weise darstellen kann. Das geschieht, indem die alte Software-Umgebung auf neuer Hardware und in neuer Systemumgebung imitiert (=emuliert) wird. Das ist ein sehr aufwändiger (Programmierungs-) Prozess. Im Gegensatz zur Migration wird dieser Aufwand allerdings nicht für jedes einzelne Objekt betrieben, sondern für Sammlungen von digitalen Objekten gleichen Typs. Die Originaldaten werden nicht verändert und es drohen daher keine Datenverluste. Vor allem für komplexe Datenarten werden Emulatoren entwickelt, zum Beispiel im Bereich der Computerspiele, wo das „Look & Feel“ eine große Rolle spielt. Allerdings ist auf die lange Sicht auch bei der Emulation zu beachten, dass die Emulatoren selbst dem technologischen Wandel unterliegen und daher von Zeit zu Zeit ebenfalls migriert (also aktualisiert) werden müssen.

Welche Erhaltungsstrategie man wählt, ist abhängig von der Art der Objekte, die langfristig erhalten werden sollen, und davon, was die wichtigsten zu bewahrenden Teile eines Objektes sind. In der Fach-Community spricht man dabei von „Signifikanten Eigenschaften“, also denjenigen Eigenschaften eines Objektes, die unbedingt erhalten werden müssen. Es lässt sich natürlich nicht exakt voraussagen, was zukünftige Nutzer erwarten, aber man kann sich dieser Frage – je nach Disziplin – annähern. So ist es wahrscheinlich, dass beispielsweise bei Kunstzeitschriften das Layout und die Farbgestaltung als bewahrenswert betrachtet werden, da es sich dabei um relevante Kriterien für eine zukünftige Analyse handeln mag. Hingegen ist bei wissenschaftlichen Texten das Layout im Gegensatz zum Inhalt möglicherweise nicht von so großer Bedeutung. Weitergehende Informationen zum Konzept der „Signifikanten Eigenschaften“ und seiner Umsetzung finden sich im nestor-Leitfaden *„Digitale Bestandserhaltung“*.²

Metadaten tragen in unterschiedlicher Art und Weise zur langfristigen Erhaltung von digitalen Objekten bei. Neben inhalterschließenden Angaben sind für die Langzeitarchivierung vor allem Daten zum Format von großer Bedeu-

2 nestor-Arbeitsgruppe Digitale Bestandserhaltung (Hrsg.) (2011): Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung. Version 1.0. nestor-materialien 15. Frankfurt am Main - 77 S. URN: urn:nbn:de:0008-2011101804 <http://nbn-resolving.de/urn:nbn:de:0008-2011101804>

tung. In welchem „Originalformat“ wurde das Objekt in das Archiv abgeliefert? In welchem Format soll es jeweils aktuell den Nutzern präsentiert werden? In welchem Format wird das Objekt archiviert? Für diese genannten Funktionen müssen nicht zwangsläufig die gleichen Formate genutzt werden. Zusätzlich werden Metadaten zu technischen Aspekten erhoben. Dabei geht es dann unter anderem darum, welche Hardware benutzt wurde, bei Bilddokumenten zum Beispiel welche Digitalkamera (Marke, Typenbezeichnung) zur Erstellung genutzt wurde. Daraus lassen sich dann später im Zweifelsfall Rückschlüsse ziehen, die für die weitere Bearbeitung notwendig sind. Gleiches gilt auch für die verwendete Software.

Nach der Durchführung einer Migration werden die Daten der dadurch vorgenommenen Änderung ebenfalls in den Metadaten dokumentiert, damit diese Schritte langfristig nachvollziehbar bleiben.

Im digitalen Bereich spielen Nutzungs- und Verwertungsrechte eine große Rolle. Diese regeln, zu welchen Bedingungen Dokumente genutzt werden dürfen, ob sie für alle frei zugänglich im Netz zur Verfügung stehen oder nur autorisierten Nutzern im Lesesaal angezeigt werden dürfen. Diese Informationen werden ebenfalls als Metadaten mit dem Objekt gespeichert.

Eines der gängigsten Metadatenformate, das im Bereich Langzeitarchivierung genutzt wird, ist „*Preservation Metadata: Implementation Strategies*“ (PREMIS)³, das an der Library of Congress entwickelt wurde. Außerdem ist „*Metadata Encoding and Transmission Standard*“ (METS)⁴ ein wichtiger Standard, ein Format zur Beschreibung von digitalen Objekten.

An der Deutschen Nationalbibliothek wurden die Langzeitarchivierungsmetadaten für elektronische Ressourcen (LMER)⁵ entwickelt.

Damit gewährleistet ist, dass digitale Objekte auffindbar bleiben, ist es notwendig, sie mit einer stabilen Adresse zu versehen. Denn oftmals verschwinden digitale Objekte im Netz und stattdessen erscheint „404 Seite nicht gefunden“ auf dem Bildschirm. URLs können sich sehr schnell ändern, daher ist es notwendig, eine stabile und eindeutige Adresse für digitale Dokumente zu vergeben. Diesen Zweck erfüllen Persistent Identifier (PI). Ähnlich wie eine ISBN eine Identifizierung gedruckter Bücher ermöglicht, fungieren PI im digitalen Bereich. Es gibt unterschiedliche Systeme, die im Prinzip aber ähnlich aufgebaut sind und eine ähnliche Funktionsweise haben. Im Verlagswesen sind Digital Object Identifier (DOI)⁶ sehr verbreitet, während im Bibliotheksbereich Uniform Resource

3 <http://www.oclc.org/research/activities/past/orprojects/pmwg/default.htm>

Deutsche Übersetzung: http://www.loc.gov/standards/premis/understanding_premis_german.pdf

4 <http://www.loc.gov/standards/mets/>

5 <http://nbn-resolving.de/urn:nbn:de:1111-2005041102>

6 <http://www.doi.org>

Names (URN)⁷ in größerem Maße genutzt werden. Das Prinzip von Persistent Identifier funktioniert folgendermaßen: Für ein digitales Dokument wird eine eindeutige Kennung generiert und diesem zugeordnet. Diese wird dann in einer Datenbank hinterlegt, zusammen mit einer oder mehreren gültigen URLs. Über einen sogenannten Resolvingprozess kann dann das Objekt angesteuert werden. Sollte eine URL nicht aufrufbar sein, wird die jeweils nächste in der Datenbank gelistete aufgerufen.

Ein weiterer wichtiger Aspekt bei der digitalen Langzeitarchivierung ist die Vertrauenswürdigkeit. Denn auch für digitale Objekte muss sichergestellt sein, dass sie authentisch, echt und unbeschädigt sind. Das bedeutet, der Betreiber eines digitalen Archivs, in dem Dokumente aufbewahrt werden, muss gewisse Kriterien erfüllen, damit Vertrauenswürdigkeit hergestellt wird. Die nestor - Arbeitsgruppe Vertrauenswürdige Archive hat – in Kooperation und Absprache mit weiteren Initiativen – eine Zusammenstellung verschiedener Kriterien erarbeitet, den Kriterienkatalog Vertrauenswürdige Archive. Dieser behandelt ganz unterschiedliche Bereiche und Prozesse, die bei der langfristigen Erhaltung digitaler Objekte eine Rolle spielen. Technische Aspekte spielen dabei eine untergeordnete Rolle, im Vordergrund stehen organisatorische Fragen und der Umgang mit den Objekten. Der nestor-Kriterienkatalog Vertrauenswürdige Archive ist Grundlage für die DIN-Norm 31644. Derzeit wird die Umsetzung des Verfahrens teilweise erprobt. Diese Zertifizierungsaktivitäten entstehen jedoch nicht separat auf nationaler Ebene, sondern sind eingebettet in einen europäischen Prozess. Das European Framework for Audit and Certification sieht ein dreistufiges Verfahren zur Zertifizierung von digitalen Archiven vor: Die erste Stufe (Basic Certification) sieht einen Selbst-Audit-Prozess des Data Seal of Approval (DSA) vor. Die zweite Stufe baut darauf auf, das bedeutet, zusätzlich zum DSA wird ein Self-Assessment entweder nach der DIN-Norm 31644 oder nach der ISO-Norm 16363 durchgeführt (Extended Certification). Als dritte Stufe wird dann, wiederum zusätzlich zur Basic Certification eine Fremdzertifizierung nach DIN 31644 oder ISO 16363 durchgeführt.⁸

Neben technischen und organisatorischen Fragen spielen rechtliche Aspekte eine große Rolle bei der digitalen Langzeitarchivierung. Unterschiedliche Bereiche und Gesetze sind davon betroffen. Kulturbewahrende Einrichtungen haben unterschiedliche Mandate in Bezug auf die Archivierung, sowohl im analogen als auch im digitalen Bereich.

Im Bibliotheksbereich gibt es inzwischen einige mit einem gesetzlichen Auftrag zur Sammlung und langfristigen Erhaltung von digitalen Objekten. Auf nationaler Ebene ist seit 2006 das Gesetz über die Deutsche Nationalbibliothek⁹

7 http://www.dnb.de/DE/Netzpublikationen/URNService/urnservice_node.html

8 <http://www.trusteddigitalrepository.eu/Site/Trusted%20Digital%20Repository.html>

9 <http://www.gesetze-im-internet.de/dnbg/index.html>

in Kraft, in welchem explizit auch die langfristige Bewahrung nicht körperlicher Medien geregelt ist. Auf Länderebene gibt es inzwischen ebenfalls einige elektronische Pflichtexemplargesetze. Damit sind die rechtlichen Grundlagen für die digitale Langzeitarchivierung geschaffen. Trotzdem ist bislang nicht abschließend geklärt, ob Veränderungen an einem Objekt, wie sie beispielsweise durch eine Migration entstehen, erlaubt sind oder nicht. Stellen diese eine Kopie dar, dann könnte es problematisch in Bezug auf das Urheberrecht werden.

Auch hinsichtlich der Bereitstellung digitaler Objekte ist die juristische Seite noch ungeklärt, beziehungsweise wird diese teilweise restriktiv gehandhabt.¹⁰ Darüber hinaus bestehen auch ganz praktische Probleme: Viele digitale Objekte sind mit einem Kopierschutz versehen oder werden mit einem Digital-Rights-Management-System verwaltet. Solche Maßnahmen verhindern eine effektive Langzeitarchivierung, weil die Objekte nicht verändert werden können. Zunehmend geraten auch Forschungsdaten aus unterschiedlichen Gründen in den Fokus der Langzeitarchivierung. Zum einen müssen Forschungsergebnisse nachvollziehbar und vor allem nachprüfbar sein, das bedeutet, sie müssen entsprechend aufbereitet werden. Zum anderen sind viele Erhebungen/Forschungsvorhaben entweder generell nicht reproduzierbar oder dies ist mit einem enormen Kostenaufwand verbunden. Daher macht es Sinn, Datensätze nachnutzen zu können. Dazu müssen nicht nur die Daten selbst, sondern auch Dokumentationen und andere Informationen, die dafür notwendig sind, zugänglich gemacht werden. In den Sozialwissenschaften kommt noch hinzu, dass viele Studien personenbezogene Daten enthalten, und in diesen Fällen sind zusätzlich noch datenschutzrechtliche Belange zu beachten.

Digitale Langzeitarchivierung betrifft sehr viele und sehr unterschiedliche Einrichtungen. Zunächst sind da die Gedächtnisorganisationen zu nennen, deren Auftrag in der Bewahrung des kulturellen Erbes besteht. Dieser ist zwar nicht immer explizit auf digitale Werke ausgedehnt, meint diese aber in einigen Einrichtungen mit. Bibliotheken, Museen und Archive müssen sich mit den Herausforderungen der digitalen Langzeitarchivierung auseinandersetzen und entsprechende Maßnahmen für ihre Sammlungen ergreifen. Dabei stehen die einzelnen Bereiche oder Communities auch vor jeweils sehr spezifischen Fragestellungen. So erfordert die langfristige Bewahrung multimedialer Kunst andere Strategien als die Langzeitarchivierung von Forschungsdaten oder die Erhaltung von Texten. Trotzdem gibt es viele Aspekte, die für alle wenn nicht gleich so doch ähnlich sind. Außerdem ist es sinnvoll, Erfahrungen auszutauschen und voneinander zu lernen.

Auch Wissenschaftlerinnen und Wissenschaftler sind in unterschiedlichen Rollen mit der Frage der langfristigen Verfügbarkeit von Daten konfrontiert: Zum einen produzieren sie Daten für ihre Forschungsvorhaben. Zum anderen

¹⁰ Siehe nestor-Stellungnahme http://files.dnb.de/nestor/berichte/nestor-Stellungnahme_AG-Recht.pdf

sind sie auch die Nutzerinnen und Nutzer von Daten. Die Frage nach der Langzeitarchivierung von Forschungsdaten wird innerhalb der wissenschaftlichen Communities zunehmend intensiver diskutiert und wirft unter anderem auch die Frage nach den jeweiligen Rollen und Verantwortlichkeiten auf. Welche Institutionen sind für die Speicherung welcher Daten zuständig? Ist jedes einzelne Forschungsinstitut für die eigenen Daten verantwortlich oder ein Institutional Repository? Gibt es ein Datenzentrum für die jeweilige Wissenschaftsdisziplin? Die Deutsche Forschungsgemeinschaft (DFG) hat in ihren Grundsätzen zur „*Sicherung der guten wissenschaftlichen Praxis*“¹¹ gefordert, dass wissenschaftliche Rohdaten zehn Jahre lang aufbewahrt werden sollen. Inzwischen gibt es Überlegungen, einen Datenmanagement-Plan als obligatorischen Bestandteil eines Förderantrags zu etablieren, ohne den keine Bewilligung erfolgen soll. Ein solcher Datenmanagement-Plan legt schon vor Beginn des Projekts fest, wie und wo die entstehenden Forschungsdaten langfristig aufbewahrt werden sollen.

Für Wirtschaftsunternehmen ist die revisionssichere Bewahrung von bestimmten Daten von zentraler Bedeutung. Zwar müssen nicht alle Daten „für immer“ archiviert werden, aber doch für festgelegte Zeiträume. Auch viele Bereiche der Politik und Verwaltung sind zunehmend digital organisiert und müssen entsprechende Maßnahmen zur Sicherung der entstehenden Daten ergreifen. Gleiches gilt für das Gesundheitswesen.

Letztendlich kommen fast alle Bereiche des Lebens zunehmend mit digitalen Daten in Berührung und müssen sich entsprechend damit auseinandersetzen, wie diese langfristig verfügbar bleiben. Es müssen nicht alle digital erzeugten Objekte für immer und ewig archiviert werden, aber es bedarf einer Festlegung, welche Daten wie lange und von wem bewahrt werden sollen.

Zusammenfassend lässt sich sagen, dass die Bestandserhaltung von analogen und digitalen Objekten sehr unterschiedliche Maßnahmen erfordert. Digitale Objekte sind an Datenträger gebunden, die spezielle Soft- und Hardware - Umgebungen benötigen, und allen gemein ist, dass sie schnell veralten. Grundsätzlich müssen mehrere Komponenten beachtet werden und verschiedene Maßnahmen zur Langzeitarchivierung eingeleitet werden. Vor allem aber handelt es sich bei der digitalen Langzeitarchivierung um einen fortlaufenden Prozess.

nestor – Kompetenznetzwerk digitale Langzeitarchivierung

Die Herausforderungen der digitalen Langzeitarchivierung sind vielfältig und können von Community zu Community variieren. Dennoch erscheint es sinnvoll, Erfahrungen und Informationen auszutauschen.

¹¹ http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf

Das Kompetenznetzwerk nestor bietet in diesem Sinne auf der einen Seite Informationen zu allen relevanten Aspekten der digitalen Langzeitarchivierung in Form von Materialien, Berichten, Checklisten etc. über seine Webseite¹² für alle Interessierten an. Auf der anderen Seite ist nestor durch seine Struktur mit einem festen Kreis an Partnern und einer großen Anzahl an Experten, die sich in den unterschiedlichen nestor-Arbeitsgruppen engagieren und austauschen, eine zuverlässige und innovative nationale Einrichtung im Bereich der digitalen Langzeitarchivierung.

nestor bündelt das vorhandene Know-how und die Kompetenzen im Bereich der digitalen Langzeitarchivierung in Deutschland. Derzeit sind 12 Einrichtungen aus den Bereichen Archiv, Museum, Bibliothek und Universität Partner von nestor.

Als community-übergreifendes Netzwerk setzt und bearbeitet nestor aktuelle und strukturbildende Themen, die gemeinsam erarbeiteten Erkenntnisse werden untereinander und mit anderen Stakeholdern geteilt. Neben technischen Entwicklungen sind besonders organisatorische Fragen im Fokus von nestor. Alle wichtigen Akteure auf nationaler Ebene sind in oder mit nestor vernetzt.

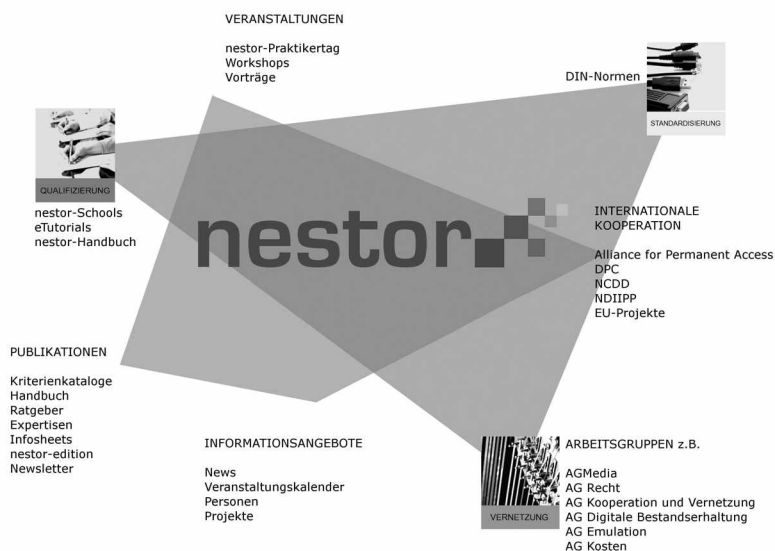


Abbildung 2: nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen für Deutschland

¹² www.langzeitarchivierung.de